	BI-4	Percent Usage Accuracy
Operator Services and Directory Assistance	OSDA-1	Mean Time to Answer
Network Performance	NP-1	Network Performance Parity
Interconnect / Unbundled Elements and Combos	IUE-1	Function Availability
	IUE-2	Timeliness of Element Performance

The Service Quality Measurements document describes the importance of each measure as an indicator of service parity. The SQM document also describes reporting dimensions that will be used to break each measure out by like factors (*e.g.*, major service group).

Why We Need to Use Statistical Tests

The Telecommunications Act of 1996 requires that ILECs provide nondiscriminatory support regardless of whether the CLEC elects to employ interconnection, services resale, or unbundled network elements as the market entry method. It is essential that CLECs and regulators be able to determine whether ILECs are meeting these parity and nondiscriminatory obligations. In order to make such a determination, the ILEC's performance for itself must be compared to the ILEC's performance in support of CLEC operations; and the results of this comparison must demonstrate that the CLEC receives no less than equal treatment compared to that the ILEC provides to its own operations. Where a direct comparison to analogous ILEC performance is not possible, the comparative standard is the level of performance that offers an efficient CLEC a meaningful opportunity to compete.

When making the comparison of ILEC results to CLEC results, it is necessary to employ comparative procedures that are based upon generally accepted statistical procedures. It is important to use statistical procedures because all of the ILEC-CLEC processes that will be measured are processes that contain some degree of randomness. Statistical procedures recognize that there is measurement variability, and assist in translating results data into useful decision-making information. A statistical approach allows for measurement variability while controlling the risk of drawing an inappropriate conclusion (*i.e,* a "type 1" or "type 2" error, discussed in the next section).

Basic Concepts and Terms

Populations and Samples

Statistical procedures will permit a determination whether the support that the ILECs provide to CLECs is indistinguishable from the support provided by the ILECs to their own customers. In statistical terms, we will determine whether two "samples", the ILEC sample and the CLEC sample, come from the same "population" of measurements.

The procedures described in this paper are based on the following assumption: *When parity is provided, the ILEC data and CLEC data can both be regarded as samples from a common population of possible outcomes.* In other words, if parity exists, the measured results for a CLEC should not be distinguishable from the measured results for the ILEC, once random variability is taken into account. Figure 1 illustrates this concept. On the right side of the figure are histograms of two samples. In this illustration, the ILEC sample contains 200 observations (data values) and the CLEC sample contains 50. Note that the two histograms are not exactly alike. This is due to sampling variation. The assumption that parity exists implies that both samples were drawn from the same population of values. If it were possible to observe this population completely, the population histogram might appear as shown on the left of the Figure. If the samples were indeed taken from this population, histograms drawn for larger and larger samples would look more and more like the population histogram. Figure 1 shows that even when parity is being provided, there will be differences between the samples due to sampling variability. Statistical tests quantify the differences between the two samples and make proper allowance for sampling variability. They assess the chance that the differences that are observed are due simply to sampling variability, if parity is being provided.
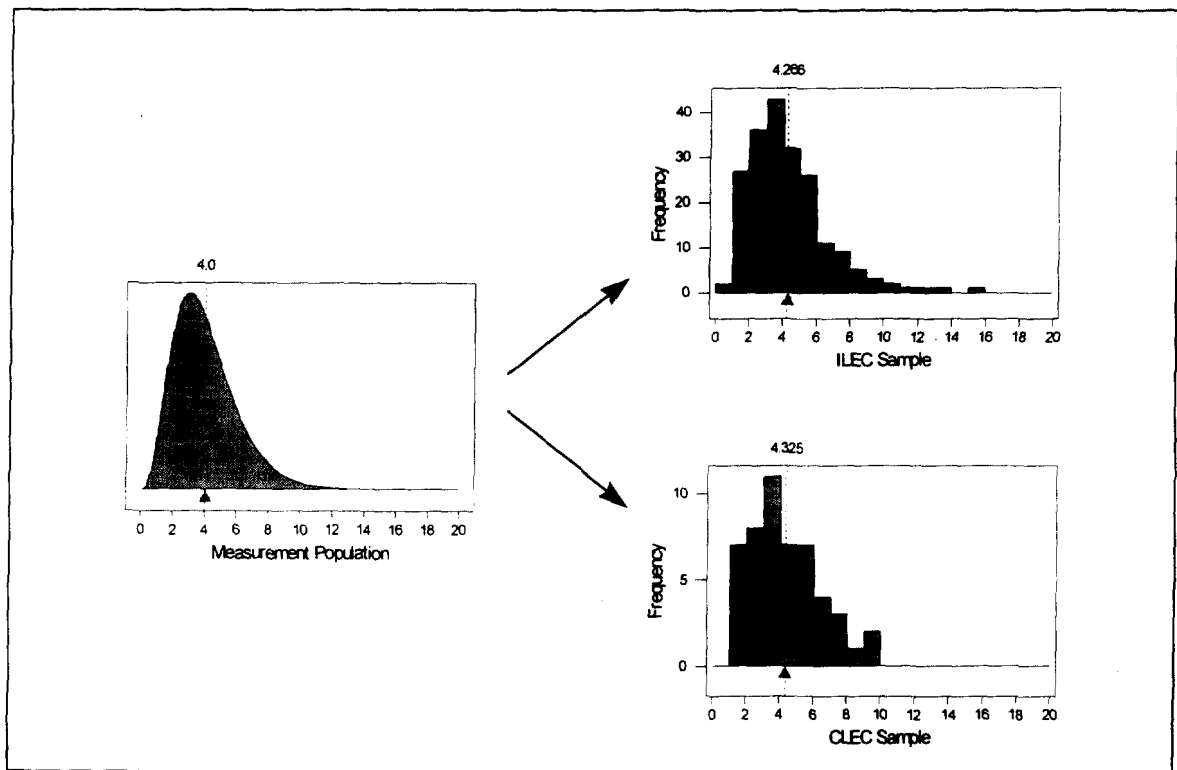

Figure 1.

Measures of Central Tendency and Spread

Often, distributions are summarized using "statistics." For the purpose of this paper, a "statistic" is simply a calculation performed on a sample set of data. Two common types of statistics are known as measures of "central tendency" and "spread."

A measure of central tendency is a summary calculation that describes the middle of the distribution in some way. The most common measure of central tendency is called the "mean" or "average" of the distribution. The mean of a sample is simply the sum of the data values divided by the sample size (number of observations). Algebraically, this calculation is expressed as

$$\bar{x} = \frac{\Sigma x}{n},$$

where x denotes a value in the sample and n denotes the sample size. The mean describes the center of the distribution in the following way: *If the histogram for a sample were a set of weights stacked on top of a flat board placed on top of a fulcrum (a "see-saw"), the mean would be the position along the board at which the board would balance.* (See Figure 1.) The mean in Figure 1 is indicated by the small triangle at approximately the value "4" on the horizontal axis.

A measure of spread is a summary calculation that describes the amount of variation in a sample. A common measure of spread is a called the "standard deviation" of the sample. The standard deviation is the typical size of a deviation of the observations in the sample from their mean value. The standard deviation is calculated by subtracting the mean value from each observation in the sample, squaring the resulting differences (so that negative and positive differences don't offset), summing the squared differences, dividing the sum by one less than the sample size, then taking the square root of the result. Algebraically, this calculation is expressed as

$$\sigma = \sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}}.$$

While the notion of mean and standard deviation exists for populations as well as samples, the mathematical definition for the mean and standard deviation for populations is beyond the scope of this paper. However, their interpretation is generally the same as for samples. In fact, for very large samples, the sample mean and sample standard deviation will be very close to the mean and standard deviation of the population from which the sample was taken.

Sampling Distribution of the Sample Mean

In Figure 1 we showed the positions of the means of the population and the two samples with triangular symbols beneath the distributions. If we sample over successive months, we will get new ILEC samples and new CLEC samples each and every month. These samples will not be

exactly like the one for the first month; each will be influenced by sampling variability in a different way. In Figure 2, we show how sets of 100 successive ILEC means and 100 successive CLEC means might appear. The ILEC means can be thought of as being drawn from a population of sample means; this population is called the "sampling distribution" of these ILEC means. This sampling distribution is completely determined by the basic population of measurements that we start with, and the number of observations in each sample. The sampling distribution has the same mean as the population.

Figure 2 illustrates two important statistical concepts:

1. The histogram of successive sample means resembles a bell-shaped curve known as the Normal Distribution. This is true even though the individual observations came from a skewed distribution.
2. The standard deviation of the distribution of sample means is much smaller than the standard deviation of the observations themselves. In fact, statistical theory establishes the fact that the standard deviation on the population of means is smaller by a factor $\sqrt{n}$, where n is the sample size. This effect can be seen in our example: the distribution of the CLEC means is twice as broad as the distribution of the ILEC means, since the ILEC sample size (200) is four times as large as the CLEC sample size (50).
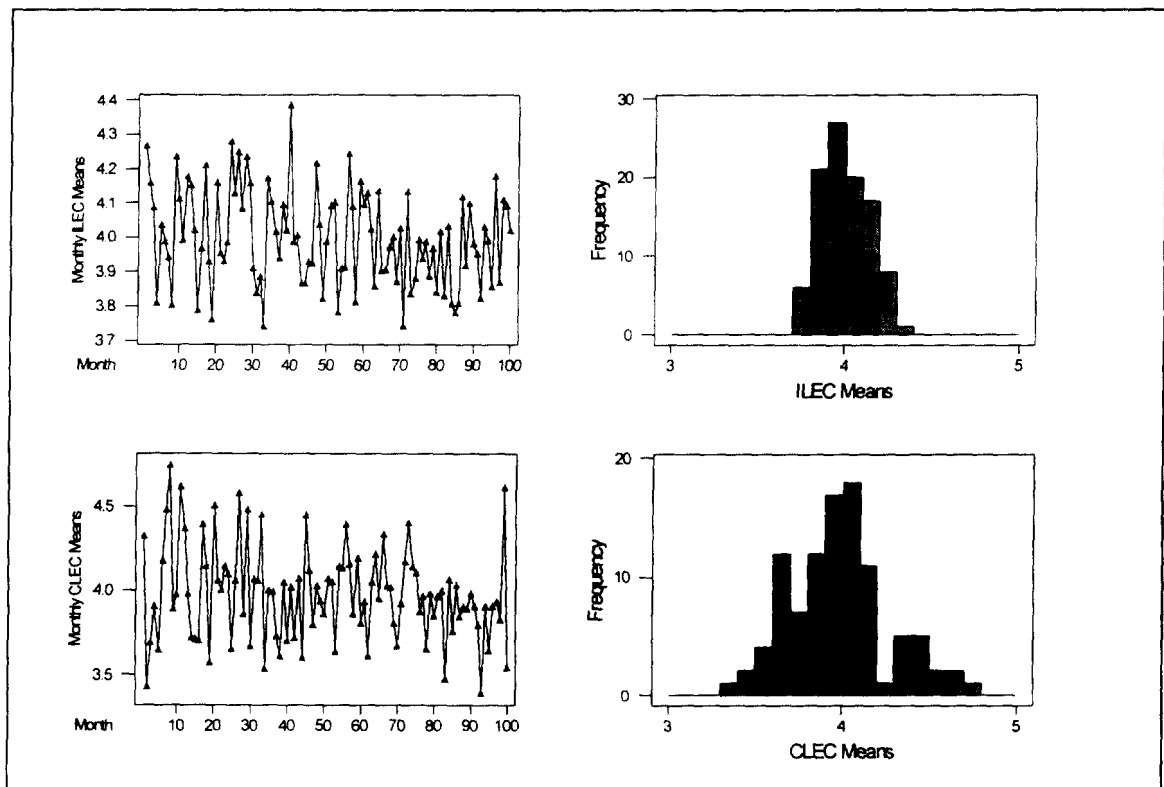

Figure 2.

It is common to call the standard deviation of the sampling distribution of a statistic the "standard error" for the statistic. We shall adopt this convention to avoid confusion between the standard deviation of the individual observations and the standard deviation (standard error) of

the statistic. The latter is generally much smaller than the former. In the case of sample means, the standard error of the mean is smaller than the standard deviation of the individual observations by a factor of $\sqrt{n}$.

The Z-test

Our objective is to compare the mean of a sample of ILEC measurements with the mean of a sample of CLEC measurements. Suppose both samples were drawn from the same population; then the difference between these two sample means (*i.e.*, $DIFF = \bar{x}_{CLEC} - \bar{x}_{ILEC}$) will have a sampling distribution which will

(i) have a mean of zero; and
(ii) have a standard error that depends on the population standard deviation and the sizes of the two samples.

Statisticians utilize an index for comparing measurement results for different samples. The index employed is a ratio of the difference in the two sample means (being compared) and the standard deviation estimated for the overall population. This ratio is known as a z-score. The z-score compares the two samples on a standard scale, making proper allowance for the sample sizes.

The computation of the difference in the two sample means is straightforward.

$$DIFF = \bar{x}_{CLEC} - \bar{x}_{ILEC}$$

The standard deviation is less intuitive. Nevertheless, statistical theory establishes the fact that

$$\sigma^2_{DIFF} = \frac{\sigma^2}{n_{CLEC}} + \frac{\sigma^2}{n_{ILEC}},$$

where σ is the standard deviation of the population from which both samples are drawn. That is, the squared standard error of the difference is the sum of the squared standard errors of the two means being compared.[1]

We do not know the true value of the population σ, because the population cannot be fully observed. However, we can estimate σ given the standard deviation of the ILEC sample (σ_{ILEC}).[2] Hence, we may estimate the standard error of the difference with

---

[1] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 370.
[2] Winkler and Hays, *Probability, Inference, and Decision*. (Holt, Rinehart and Winston: New York), p. 338.

$$\sigma_{DIFF} = \sqrt{\frac{\sigma_{ILEC}^2}{n_{CLEC}} + \frac{\sigma_{ILEC}^2}{n_{ILEC}}} = \sqrt{\sigma_{ILEC}^2 \left[\frac{1}{n_{CLEC}} + \frac{1}{n_{ILEC}} \right]}$$

If we then divide the difference between the two sample means by this estimate of the standard deviation of this difference, we get what is called a "z-score".

$$z = \frac{DIFF}{\sigma_{DIFF}}$$

Because we assumed that both samples were in fact drawn from the same population, this z-score has a sampling distribution that is very nearly Standard Normal, *i.e.*, having a mean of zero and a standard error of one. Thus, the z-score will lie between ± 1 in about 68% of cases, will lie between ± 2 in about 95% of cases, and will lie between ± 3 in about 99.7% of cases, always assuming that both samples come from the same population. Therefore, one possible procedure for checking whether both samples come from the same population is to compare the z-score with some cut-off value, perhaps +3. For comparisons where the values of z exceed the cutoff value, you reject the assumption of parity as not proven by the measured results. This is an example of a statistical test procedure. It is a formal rule of procedure, where we start with raw data (here two samples, ILEC measurements and CLEC measurements), and arrive at a decision, either "conformity" or" violation".

Type 1 Errors and Type 2 Errors

Each statistical test has two important properties. The first is the probability that the test will determine that a problem exists when in fact there is none. Such a mistaken conclusion is called a type one error. In the case of testing for parity, a type one error is the mistake of charging the ILEC with a parity violation when they may not be acting in a discriminatory manner. The second property is the probability that the test procedure will not identify a parity violation when one does exist. The mistake of not identifying parity violation when the ILEC is providing discriminatory service is called a type two error. A balanced test is, therefore, required.

From the ILEC perspective, the statistical test procedure will be unacceptable if it has a high probability of type one errors. From the CLEC perspective, the test procedure will be unacceptable if it has a high probability of type two errors.

Very many test procedures are available, all having the same probability of type one error. However the probability of a type two error depends on the particular kind of violation that occurs. For small departures from parity, the probability of detecting the violation will be small. However, different test procedures will have different type two error probabilities. Some test procedures will have small type two error when the CLEC mean is larger than the ILEC mean, even if the CLEC standard deviation is the same as the ILEC standard deviation, while other procedures will be sensitive to differences in standard deviation, even if the means are equal.

Our proposals below are designed to have small type two error when the CLEC mean exceeds the ILEC mean, whether or not the two variances are equal.

Tests of Proportions and Rates

When our measurements are proportions (*e.g.* percent orders completed on time) rather than measurements on a scale, there are some simplifications. We can think of the "population" as being analogous to an urn filled with balls, each labeled either 0(failure) or 1(success). In this population, the fraction of 1's is some "population proportion". Making an observation corresponds to drawing a single ball from this urn. Each month, the ILEC makes some number of observations, and reports the ratio of failures or successes to the total number of observations; the ILEC does the same does the same for the CLEC. The situation is very similar to that discussed above; however, rather than a wide range of possible result values, we simply have 0's (failures) and 1's (successes). The "sample mean" becomes the "observed proportion", and this will have a sampling distribution just as before. The novelty of the situation is that now the population standard deviation is a known function of the population proportion[3]; if the population proportion is p, the population standard deviation is $\sqrt{p(1-p)}$, with similar simplifications in all the other formulas.

There is a similar simplification when the observations are of rates, *e.g.*, number of troubles per 100 lines. The formulas appear below.

Proposed Test Procedures

Applying the Appropriate Test

Three z-tests will be described in this section: the "Test for Parity in Means", the "Test for Parity in Rates", and the "Test for Parity in Proportions". For each LCUG Service Quality Measurement (SQM), one or more of these parity tests will apply. The following chart is a guide that matches each SQM with the appropriate test.

Preordering Response Interval (PO-1)	Mean
Avg. Order Completion Interval (OP-1)	Mean
% Orders Completed On Time (OP-2)	Proportion
% Order (Provisioning) Accuracy (OP-3)	Proportion
Order Reject Interval (OP-4)	Mean
Firm Order Confirmation Interval (OP-5)	Mean
Mean Jeopardy Interval (OP-6)	Mean

---

[3] Winkler and Hays, *Probability, Inference, and Decision.* (Holt, Rinehart and Winston: New York), p. 212.

Completion Notice Interval (OP-7)	Mean
Percent Jeopardies Returned (OP-8)	Proportion
Held Order Interval (OP-9)	Mean
% Orders Held ≥ 90 Days (OP-10)	Proportion
% Orders Held ≥ 15 Days (OP-11)	Proportion
Time To Restore (MR-1)	Mean
Repeat Trouble Rate (MR-2)	Proportion
Frequency of Troubles (MR-3)	Rate
Estimated Time To Restore (MR-4)	Proportion
System Availability (GE-1)	Proportion
Center Speed of Answer (GE-2)	Mean
Call Abandonment Rate (GE-3)	Proportion
Mean Time to Deliver Usage Records (BI-1)	Mean
Mean Time to Deliver Invoices (BI-2)	Mean
Percent Invoice Accuracy (BI-3)	Proportion
Percent Usage Accuracy (BI-4)	Proportion
OS/DA Speed of Answer (OS/DA-1)	Mean
Network Performance (NP-1)	Mean, Proportion
Availability of Network Elements (IUE-1)	Mean, Proportion
Performance of Network Elements (IUE-2)	Mean, Proportion

Test for Parity in Means

Several of the measurements in the LCUG SQM document are averages (*i.e.*, means) of certain process results. The statistical procedure for testing for parity in ILEC and CLEC means is described below:

1. Calculate for each sample the number of measurements (n_{ILEC} and n_{CLEC}), the sample means ($\bar{x}_{ILEC}$ and $\bar{x}_{CLEC}$), and the sample standard deviations (σ_{ILEC} and σ_{CLEC}).

2. Calculate the difference between the two sample means; if *larger* CLEC mean indicates possible violation of parity, use $DIFF = \bar{x}_{CLEC} - \bar{x}_{ILEC}$, otherwise reverse the order of the CLEC mean and the ILEC mean.

3. To determine a suitable scale on which to measure this difference, we use an estimate of the population variance based on the ILEC sample, adjusted for the sized of the two samples: this gives the standard error of the difference between the means as

$$\sigma_{DIFF} = \sqrt{\sigma_{ILEC}^2 \left[\frac{1}{n_{CLEC}} + \frac{1}{n_{ILEC}} \right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{DIFF}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example (double-click to edit)

Test for Parity in Proportions

Several of the measurements in the LCUG SQM document are proportions derived from certain counts. The statistical procedure for testing for parity in ILEC and CLEC proportions is described below. It is the same as that for means, except that we do not need to estimate the ILEC variance separately.

1. Calculate for each sample sample sizes (n_{ILEC} and n_{CLEC}), and the sample proportions (p_{ILEC} and p_{CLEC}).

2. Calculate the difference between the two sample means; if *larger* CLEC proportion indicates worse performance, use $DIFF = p_{CLEC} - p_{ILEC}$, otherwise reverse the order of the ILEC and CLEC proportions.

3. Calculate an estimate of the *standard error for the difference* in the two proportions according to the formula

$$\sigma_{DIFF} = \sqrt{p_{ILEC}(1 - p_{ILEC})\left[\frac{1}{n_{CLEC}} + \frac{1}{n_{ILEC}}\right]}$$

4. Hence compute the test statistic

$$z = \frac{DIFF}{\sigma_{DIFF}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example (double-click to edit)

Test for Parity in Rates

A rate is a ratio of two counts, *num/denom*. An example of this is the trouble rate experience for POTS. The procedure for analyzing measurements results that are rates is very similar to that for proportions.

1. Calculate the numerator and the denominator counts for both ILEC and CLEC, and hence the two rates $r_{ILEC} = num_{ILEC}/denom_{ILEC}$ and $r_{CLEC} = num_{CLEC}/denom_{CLEC}$.

2. Calculate the difference between the two sample rates; if *larger* CLEC rate indicates worse performance, use $DIFF = r_{CLEC} - r_{ILEC}$, otherwise take the negative of this.

3. Calculate an estimate of the *standard error for the difference* in the two rates according to the formula

$$\sigma_{DIFF} = \sqrt{r_{ILEC}\left[\frac{1}{denom_{CLEC}} + \frac{1}{denom_{ILEC}}\right]}$$

4. Compute the test statistic

$$z = \frac{DIFF}{\sigma_{DIFF}}$$

5. Determine a critical value c so that the type one error is suitably small.

6. Declare the means to be in violation of parity if $z > c$.

Example (double-click to edit)

13